REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2002	Book Chapter - Clinical Chemistry

4. TITLE AND SUBTITLE
Quality Assurance

5. FUNDING NUMBERS

6. AUTHOR(S)
Mark D. Kellogg

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute of Environmental Medicine
Kansas Street
Natick, MA 01760-5007

8. PERFORMING ORGANIZATION REPORT NUMBER
M-01/10

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.
Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Quality assurance is often used to refer to all of the processes a laboratory takes to ensure quality in test results. However, assurance is more properly defined as a pledge, or guaranty, a self-confidence. Thus, using this definition quality assurance is the result of the processes focused on achieving quality. That is, by using planning, assessment, and monitoring tools, we can pledge, guaranty, and assure the quality of the products laboratories produce. The first part of this chapter will address the quality management framework that is integral to the implementation of an effective quality control program. This will be followed by a more detailed description of the quality control monitoring process and the statistics used in this process. This chapter will utilize the terms and concepts proposed by Westgard, Burnett, and Bowers as part of their framework for managing the quality of laboratory tests.

14. SUBJECT TERMS
Quality assurance, quality control, statistics

15. NUMBER OF PAGES
23

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)
01-01-2002		01-01-2002 to 31-12-2002

4. TITLE AND SUBTITLE
Quality Assurance
Unclassified

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kellogg, Mark D. ;

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME AND ADDRESS
USARIEM
42 Kansas St.
Natick, MA01760

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS
,

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES

14. ABSTRACT
See report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Same as Report (SAR)	23	Rice, Teresa teresa.rice@na.amedd.army.mil
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number DSN

CHAPTER 4

Quality Assurance

Mark D. Kellogg

Quality assurance is often used to refer to all of the processes a laboratory takes to ensure quality in test results. However, assurance is more properly defined as a pledge, or guaranty, a self-confidence. Thus, using this definition quality assurance is the ***result*** of the processes focused on achieving quality. That is, by using planning, assessment, and monitoring tools, we can pledge, guaranty, and assure the quality of the products laboratories produce.

The first part of this chapter will address the quality management framework that is integral to the implementation of an effective quality control program. This will be followed by a more detailed description of the quality control monitoring process and the statistics used in this process. This chapter will utilize the terms and concepts proposed by Westgard, Burnett, and Bowers as part of their framework for managing the quality of laboratory tests.[1]

THE QUALITY MANAGEMENT FRAMEWORK

The first step in achieving quality laboratory processes requires a definition of the goals and objectives and the establishment of quality requirements to satisfy the customer. Without these definitions, there is no way to measure whether acceptable quality is being achieved. Imagine having your clinical chemistry instructor start a lecture stating, "Today you will learn" and then describing the process of changing tires on semitrailers. By the end of class, you will have learned something, but does it satisfy your need as a student in clinical chemistry? Was there any relevant quality in the instruction? This example shows the importance of having well-defined goals and objectives.

The next step in the framework is **quality planning.** This encompasses the creation, selection, and validation of the methods and processes used in the laboratory. What are the best available methods? Do they meet the quality requirements? The selection of quality control procedures is an example of quality planning and will be addressed later in this chapter. Remember, the plan is what gets you to the desired quality. As an example, imagine preparing for a long journey. You will need to decide how you will get there. Will you use a plane, car, train? What routes will you travel to get there? It is very difficult to accomplish the goal without a plan.

Quality laboratory processes (QLP) result from the quality planning and describe the means by which work is conducted in the laboratory. Personnel policies, standard operating procedures (SOPs), specimen-collection guidelines all belong to the QLP domain. These should represent the best way for your laboratory to get the work done. But no process is perfect and the remaining components of the management framework serve to monitor these processes and provide correction if they deviate from the established quality requirements.

Quality assessment (QA) in the clinical laboratory includes all actions a laboratory takes to measure and monitor performance of the laboratory processes. QA includes verification of the quality of sample collection, sample processing, reporting of results, and interpretation of the final report by the physician. In addition, quality assessment should address the prompt reporting and better use of test results as well as the competency and adequacy of laboratory staff.

If QA activities determine that quality requirements are not being achieved, then **quality improvement** (QI) processes must determine the cause of the problem and provide input to further quality planning to eliminate the problem. A good example would be the monitoring of hemolyzed blood samples arriving in the laboratory. A QA program may keep track of these and create a warning if more than X specimens arrive in a defined period. If this warning is given, an investigation (a QI process) might find that the specimens all came from a specific lot of evacuated tubes or resulted from incomplete training of specimen-collection personnel. Once the problem is identified, then quality planning can be used to create measures for assuring evacuated tube quality or additional training can be designed for specimen-collection personnel.

Quality improvement is more than just simply changing something to fix a problem, but a thorough assessment of what the problem is and then the use of quality planning to modify or create quality laboratory processes to eliminate the problem. Dr. Westgard provides an excellent example from his essay on total quality management.[2] In his example he describes how a physician who has ordered a stat or emergency test calls for the answer before the specimen has arrived in the laboratory. The physician insists that the laboratory needs to do testing faster. However, performing the test faster does not solve

The views, opinions, and/or findings contained in this chapter are those of the author and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

the problem, which most likely is due to problems in the collection and transport of specimen to the laboratory. Listening to the customer (physician) does not mean doing what the customer says. It is important to use your knowledge of the testing process to analyze where problems might exist and investigate to determine the root cause.

Quality control (QC) provides the tools to detect problems early and prevent errors from exceeding established quality requirements. QC is somewhat narrowly defined as the monitoring of work processes, detecting problems and making corrections prior to delivery of products or services.[1] This contrasts to the broader measures of QA described earlier that detect errors after they have occurred but before they reach a defined "critical" level. Typically, QC is applied to the analytical process (the actual measurement of analytes in a specimen) and not the pre-analytical or post-analytical phases. Statistical process control, or statistical quality control, is the major procedure for monitoring the analytical performance of laboratory methods.

QC will be the focus of the remainder of this chapter. Readers are highly encouraged to learn more about the total quality management approach to laboratory testing and to read the works of Westgard and others on quality control. The World Wide Web (WWW) provides an amazing abundance of resources, and in particular Dr. Westgard has an excellent web site at www.westgard.com.

QUALITY CONTROL MONITORING

Selection of Quality Control Materials

Before discussing the criteria for selection of control material, it is appropriate to make a distinction between a standard and a control. A **standard** is a solution that contains a known amount of an analyte and is used to calibrate an assay method. A standard has one assigned value; for example, it contains 100 mg/dL of glucose. Not 100 mg/dL ± X amount, but exactly 100 mg/dL. If we know that we are putting 100 mg/dL of glucose into the instrument, the result should be 100 mg/dL. If the result is not 100 mg/dL, then the system is adjusted (calibrated) to produce the desired 100 mg/dL value. This adjustment may be performed by you or automatically by the microprocessors controlling the instrument. Assay results are then calculated from calibration settings. **Controls,** on the other hand, are used to monitor the performance of an assay method once it has been calibrated. Controls are run alongside patient samples and results are calculated from calibration data in the same manner that patient results are calculated. The controls are used to make sure that the measurement process is stable (the same as it was when calibrated) and if the results produced are acceptable.

A control should have the same **matrix** or closely mimic the characteristics of the patient sample that is being tested. Matrix refers to all of the characteristics of the sample. In broad terms the matrix can be defined as serum, plasma, or urine. However, controls usually have additional material added as preservatives, or may be entirely synthetic. When these additional materials or components of synthetic controls do not act like a biological sample when tested, we call this **matrix interference.** Because of this interference, it is important that a test designed to be performed on serum have control material that is as similar to serum as possible. This can be accomplished by pooling actual samples of human serum or animal serum, lyophilizing or dehydrating serum samples, or using an artificial protein-based solution that has the same physical characteristics as serum. These control samples are usually purchased from commercial vendors, often from the manufacturer of the instrument or reagents that you are using.

Every QC system should include at least two levels of controls with analyte concentrations focused at medical-decision levels. The importance of this is illustrated in the following example.

A physician receiving an abnormal creatinine result on a patient may make two very different medical decisions based on the magnitude of the abnormal value:

1. For some patients, the physician may decide the value is elevated enough from normal to diagnose kidney dysfunction.
2. For other patients, the physician may decide the result is so elevated that aggressive medical intervention such as dialysis is necessary.

The laboratory that has this physician as a client could have one control focused at or near the value that distinguishes normal from abnormal and the other control focused at a value that would indicate a need for extensive medical intervention.

Many chemistry laboratories use three levels of controls with values focused at low normal, high normal, or very abnormal concentrations. An example that illustrates the benefit of using three levels of controls is shown by the physician who treats diabetic patients.

This physician might make three levels of medical decisions based on abnormal glucose values. The physician may decide that:

1. a patient has a value elevated enough to have his or her condition diagnosed as diabetes.
2. a patient with a previously diagnosed condition of diabetes has a value so elevated that the patient requires immediate insulin therapy.
3. the diabetic patient has a value low enough to be near insulin shock.

Thus, a laboratory with three levels of controls could adequately ensure quality results at all three decision points.

It is advantageous to have sufficient amounts of the same lot number of control material to last at least a year. This is because whenever a new lot of control is put into place, a lab-

oratory must reestablish target ranges for the control pool. This process, if performed correctly, takes approximately a month and would be inconvenient to perform more than once a year. In addition to the sufficient amounts to last at least a year, it is important that the control material remain stable for that period of time.

Controls should also be available in aliquots convenient for daily use. Once a control vial is opened, it is subject to evaporation and thus a change in concentration could easily occur. Commercial controls are usually available in 1-mL, 5-mL, or 10-mL aliquot vials.

Statistical Procedures

Quality control samples can be used to monitor both quantitative (produces a numerical result) and qualitative (produces a nonnumerical or positive or negative result) tests. Since most chemistry procedures are quantitative, the focus of this section will be on the statistical procedures for evaluating QC results that are numerical or quantitative.

Once controls have been selected, the next step is to assay the selected controls for a period of time so that sufficient data points are generated to establish the target range. When a control is put into operation and run alongside patient samples, the value of the control is monitored against the target range. This monitoring process is the basis for determining acceptance or rejection of patient sample results.

The usual procedure for establishing a target range is to assay each control at least once per day or once per shift for a full month. A good rule of thumb is that data sets should contain at least 20 points if the statistics calculated on the data are to be considered valid. Therefore, the time allowed to collect data to establish a control's target range should be sufficient to produce at least 20 data points. In addition, the control should be assayed over an extended period to ensure that it is subjected to changes in the laboratory environment that may occur with different operators or at different times of day.

In an ideal world, repeated sampling of a control should produce the same result each time. However, the world is not perfect and unfortunately neither is the laboratory. There will always be a certain amount of variability in repeated measurements. This variability is affected by operator technique, the laboratorian, and by the inherent variability in the assay method (the instrument or procedure used to run the test). By measuring the control sample over 20 different time points, we hope to measure this inherent variation and use this information to decide if the process changes later.

Data from repeated measurements will have a distribution or spread in the values that reflect how easy it is to repeat the measurement and obtain the same value. We can visually represent this variability in repeat measurements with frequency plots. To design a frequency plot from repeat measurements of controls, the values obtained are plotted on the x-axis, and the number of times a value is obtained is plotted on the y-axis. The resulting plot takes the shape of a curve, with the area under the curve representing the number of data points at each control value.

Example 4–1

Laboratory A assays a control for glucose 2 to 3 times per day for 30 days. The laboratory then lists all assay values in descending order and counts the number of times each value was obtained. The following table lists the laboratory values for the control and the frequency each value was obtained during the 30 day period.

Value (mg/dL)	Frequency
106	1
105	1
104	1
103	2
102	3
101	4
100	3
99	2
98	1
97	1
96	1

The laboratory labels the x-axis of a frequency plot with the range of values obtained on the control (96 to 106) and the y-axis of the plot with the minimum to maximum number of times the values were obtained (1 to 4). As seen in Figure 4–1, once the points were plotted and the curve drawn, the area under the curve gave a visual representation of the frequency of data points at each value.

As can be seen from the curve in Figure 4–1, the value that is in the middle of the range of values is also the value that is repeated most often. Therefore, the peak of the curve is at this value. Conversely, the values that have only one data point or are rarely repeated are also the values that are the lowest and the highest numbers in the data set. Thus, these values are found in the tails of the curve.

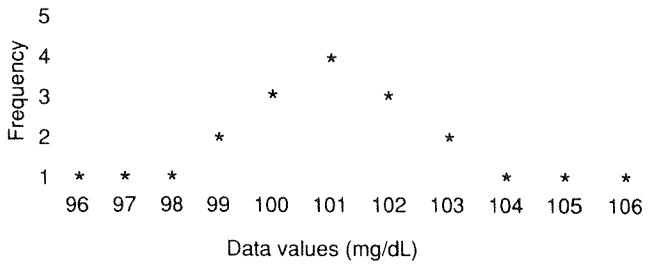

FIGURE 4-1

A frequency plot.

CENTRAL TENDENCY AND NORMAL DISTRIBUTION

Although it is impossible to obtain the same value for repeated measurements of a control, it is desirable that the majority of data points be nearly the same. This concept of clustering of data points about one value is referred to as central tendency. The frequency plot that results from data that exhibit central tendency has a peak that represents the value that is repeated most often. If the data points to the right of the peak (positive direction) are about equal to the data points to the left of the peak (negative direction), the data are said to have a normal distribution about the point of central tendency. The curve that results from plotting data that have a normal distribution is bell-shaped, with most values having frequencies in the top portion of the curve and few values having frequencies in the tails of the curve. These characteristics give the curve its spreading symmetrical or bell shape as shown in Figure 4–2.

Although plotting the data and observing the shape of the curve will indicate whether the data display central tendency and normal distribution, the interpretation of the curve is left to the observer. Statistical analysis of the curve will produce a mathematical picture of its shape that is less open to interpretation.

The statistical parameters used to measure central tendency are the mean, median, and mode. The **mean** is defined as the average of all the data points or values. The formula for calculating the mean is:

$$\bar{x} = \frac{\sum x_i}{n}$$

where $\bar{x}$ = statistical designation for the mean
$\sum$ = symbol for "sum of"
x_i = each data point or value
n = the number of data points or values

The **median** is defined as the middle data point observed once the data are arranged in descending or ascending order. It is calculated by listing the data points in numerical order (including repeat data points) and selecting the middle value. For example, in a data set of 19 data points arranged in numerical order from 1 to 19, the value listed at position 10 is the median. Nine numbers are below this value (positions 1 to 9) and 9 numbers are above this value (positions 11 to 19). If the data set contains an even number of data points, the values at the two middle positions are averaged to obtain the median. For example, in a data set containing 20 values, the value at position 10 and the value at position 11 are averaged and that average becomes the median.

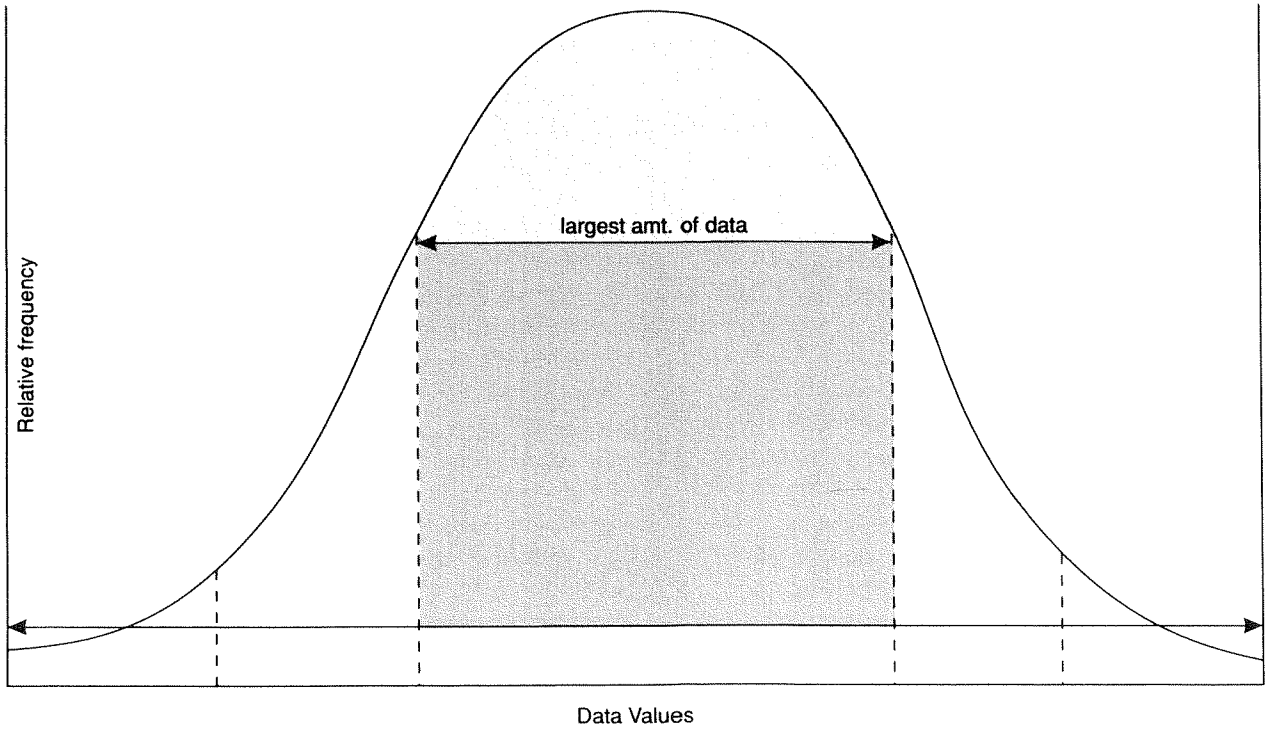

FIGURE 4-2

Normal distribution curve.

The **mode** is the value that occurs with the greatest frequency. In Example 4–1, the mode was 101 because it was the value repeated most often.

When data have a normal distribution, the mean, median, and mode are about the same. In other words, the central tendency about which the data are clustered is not only the value repeated most often but is also the middle value of the data set as well as the average of all values. This mathematically confirms that the data are clustered about the point of central tendency, the values above the peak are about equal to those below and the majority of values have frequencies near the peak.

In the example below, the calculated mean, median, and mode are essentially the same, and thus the data are normally distributed. Some data sets will have central tendency but not normal distribution because the spread of points about the center is not symmetrical. The curve from this type of data set is said to have a skewed distribution and is asymmetrical because there are more data points on one side of the peak than on the other.

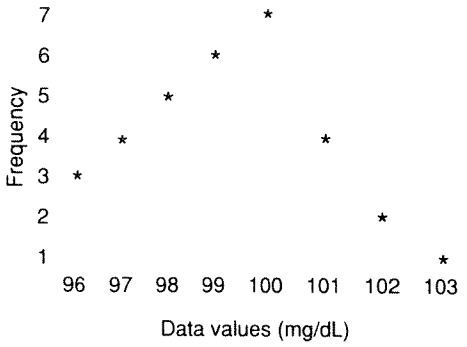

FIGURE 4-3

This frequency plot is from data that are skewed towards the left side of the peak.

Example 4–2

The first two columns in the table below list the data from Example 4–1 by date in the order in which the control was sampled. The last two columns of the table list each data point, including all repeats, by its position in descending order.

Date	Value	Position	Descending Order
6-1	103	1	106
6-2	99	2	105
6-3	100	3	104
6-4	101	4	103
6-5	98	5	103
6-8	105	6	102
6-9	102	7	102
6-10	101	8	102
6-11	103	9	101
6-12	99	10	101}middle
6-15	102	11	101}middle
6-16	97	12	101
6-17	101	13	100
6-18	96	14	100
6-19	102	15	100
6-22	104	16	99
6-23	106	17	99
6-24	100	18	98
6-25	101	19	97
6-26	100	20	96

SUM = 2020

The mean, median, and mode for the data in Example 4–2 are as follows:

Mean = 101 (2020 divided by 20 data points)
Median = 101 (indicated in last column of example 4–2)
Mode = 101

The calculated mean, median, and mode for the data represented by the frequency plot in Figure 4–3 are as follows:

Mean = 98.9
Mode = 100
Median = 99

Since the three measures of central tendency are not close enough to be considered equal, mathematical calculations indicate that the data are not normally distributed.

There are some instances in which data sets may have two points of central tendency. As shown in Figure 4–4, the curve resulting from these data has two peaks and is thus termed bimodal. The calculated mean, median, and the mode for these data are as follows:

Mean = 101.2
Median = 101
Mode = 99

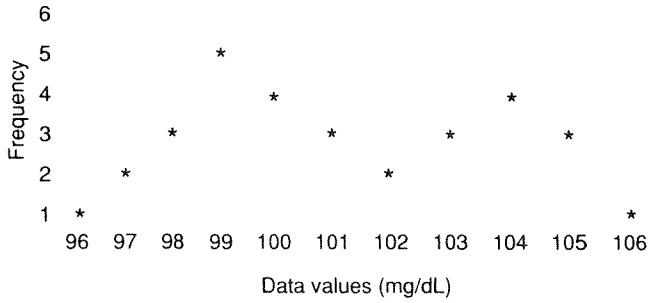

FIGURE 4-4

Curve resulting from data set with a bimodal distribution.

These calculations are not the same and therefore the curve is not normally distributed.

Since the goal for repeat measurements of controls is that all values be about the same, it is critical that their frequency plots exhibit true central tendency with very little bias on either side of the peak. Therefore, the mean, median, and mode should always be calculated to confirm that the data are normally distributed before attempting to establish a target range.

MEASURES OF DISPERSION

Once it has been determined that the data for repeat control measurements are normally distributed, it is important to consider the dispersion or the spread of the data within the distribution about the mean. The spread of data around a central tendency can be tight or broad, depending on the ability of a method or an operator to repeat the same value for a control each time it is sampled.

Both the curves in Figure 4–5 and Figure 4–6 have central tendency. The mean, median, and mode for each are about the same and thus the curves are normally distributed. However, the curve in Figure 4–5 has more points at or about the mean than the curve in Figure 4–6, and thus the data from Figure 4–5 are closer to that "perfect world" of obtaining the same value for each repeat sampling of the control. Since our goal is to have repeat control values that are as close to one value as possible, it is desirable that control data have a slim distribution about the mean. We assume that if we are successful at repeating the value for a control, we will be just as successful at repeating a value for a patient sample. Thus the ability to successfully repeat measurements and obtain the same value gives credibility to lab results.

As with determining central tendency of data, it is possible to draw a frequency plot and observe the curve for tight or broad distribution of points about the mean. Since this depends on interpretation by the observer, which can be subjective, there are statistical parameters that will give us a mathematical picture of the spread of the data about the mean. Those parameters are the range, variance, standard deviation, and coefficient of variation.

Range is defined as the difference between the highest and lowest value in a data set. To calculate the range, arrange the data set in order and subtract the lowest value from the highest value. Looking at the data in Example 4–1, the highest value in the data set is 106 and the lowest value is 96. Therefore the range of the data is 10. If we add the units of mg/dL to the data, we can now say that repeated measurements of the control pool covered a range of 10 mg/dL.

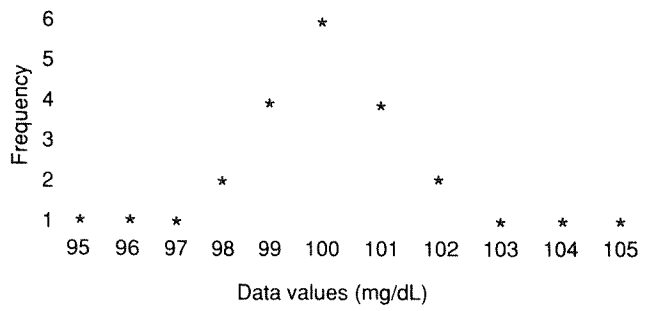

FIGURE 4-5

Curve for tight distribution of points about the mean.

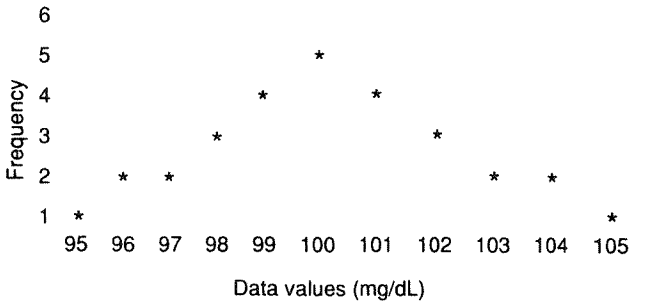

FIGURE 4-6

Curve for broad distribution of points about the mean.

Although the range gives an indication of the dispersion of data points, it is not clear how the data within the range vary about the mean. **Variance** is defined as the measure of the average squared distance of data points from the mean.

The formula for calculating variance is:

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

where S^2 = statistical symbol for variance
$\sum$ = symbol for "the sum of"
$(x_i - \bar{x})^2$ = (each data point minus the mean), squared
$n - 1$ = number of data points minus one degree of freedom

As you can see from the formula, the calculation of the variance is actually the calculation of an average. The data set that is averaged is the squared differences of data points from the mean. As with any average, the variance is calculated by summing members of a data set and dividing by the number of data points. In the variance formula, the total number of data points are adjusted for a loss in degrees of freedom.

Degrees of freedom are defined as the number of independent data points that are contained in a data set. Suppose you are told to select any 5 numbers and calculate their average. You have complete freedom in the selection of numbers since no restriction has been placed on the data set. You, therefore, have 5 independent numbers in the set. However, if you are told to take the average of any 5 numbers whose sum is 20, there is now a restriction placed on the data set and you have only 4 independent numbers. Once you select 4 numbers, the fifth number is set. For example, if you select 2, 3, 5, and 6 for the first 4 numbers, the last number must be 4 to obtain the sum of 20. Therefore, you have lost one degree of freedom.

Example 4–3

In the following example, the variance is calculated on the data from Example 4–1. The mean of the data was calculated previously as 101.

x	$x - \bar{x}$	$(x - \bar{x})^2$
	$(\bar{x} - 101)$	
103	2	4
99	−2	4
100	−1	1
101	0	0
98	−3	9
105	4	16
102	1	1
101	0	0
103	2	4
99	−2	4
102	1	1
97	−4	16
101	0	0
96	−5	25
102	1	1
104	3	9
106	5	25
100	−1	1
101	0	0
100	−1	1
SUM = 2023		SUM = 122

n = 20

When we calculate the variance of a data set, we have previously calculated the mean and will use this in the variance formula. Therefore, we have lost one degree of freedom in the data and the sum of the squared differences must be divided by "n − 1" instead of "n".

The calculations for variance in Example 4–3 are as follows:

Sum of squared differences = 122

n − 1 = 19

Variance = 6.4 (122 divided by 19)

For the purpose of ease in demonstration, the mean and the differences from the mean have been expressed as whole numbers. The mean for these data is actually a decimal that should make the differences decimals. Also, actual statistical calculations, such as those performed on calculators, will retain at least 2 decimal places beyond the original number throughout all steps in the calculation. Therefore, this data set should have 3 decimal places in the numbers listed under the squared differences column. Consequently, if these data are subjected to calculator manipulations, the final standard deviation (SD) will be slightly different.

If the original units in the data were mg/dL, the unit for the final variance calculation would be mg^2/dL^2. Although we have used a calculation that should give us valuable information about the spread of data, the final answer is of little use because the units are difficult to interpret. However, it is necessary to use the squared differences instead of the actual differences to compensate for the positive and negative variability about the mean. By squaring the differences, the resulting numbers are positive integers and represent the distance from the mean without taking direction into account.

Standard deviation is defined as the square root of the variance. The SD is merely a mathematical manipulation of the variance that converts it to a more usable statistic.

The formula to calculate SD is:

$$SD = \sqrt{\frac{(x_i - \bar{x})^2}{n - 1}}$$

As you can see, the formula for SD is merely the square root of the formula for variance. The advantage to this calculation is that the units for SD are the same as the units in the original data set.

From Example 4–3:

Variance = 6.4
SD = 2.5 (the square root of 6.25)

The SD is a measurement statistic that describes the average distance each data point in a normal distribution is from the mean. The SD unit of measure covers about one sixth the total distance of the x-axis on a normal distribution curve. If the mean of the data is the middle point on the x-axis, then 3 SD units lie on the right side of the mean and 3 SD units lie on the left side of the mean.

As shown in Figure 4–7, specific areas under a normal distribution curve are bound by the SD units on the x-axis. As you will recall, most of the data in a normal distribution lie close to the mean. Therefore, the area of the curve cut by the SD units immediately to the right and left of the mean would contain the largest number of data points. In fact, the SD unit immediately to the right and the SD unit immediately to the left of the mean each contain about 34.1% of the data. As we progress down the x-axis in both directions, the next SD unit takes in about 13.65% of the data on each side of the mean. The final SD units on each end of the x-axis contain 2.1% of the data. The remaining 0.3% of the data theoretically spreads out to infinity.

Confidence intervals are defined as the limits between which we expect a specified proportion of a population to lie. In statistical analysis of repeat measurements that follow a normal distribution, confidence limits refer to the percentage of data contained within intervals that includes the mean and

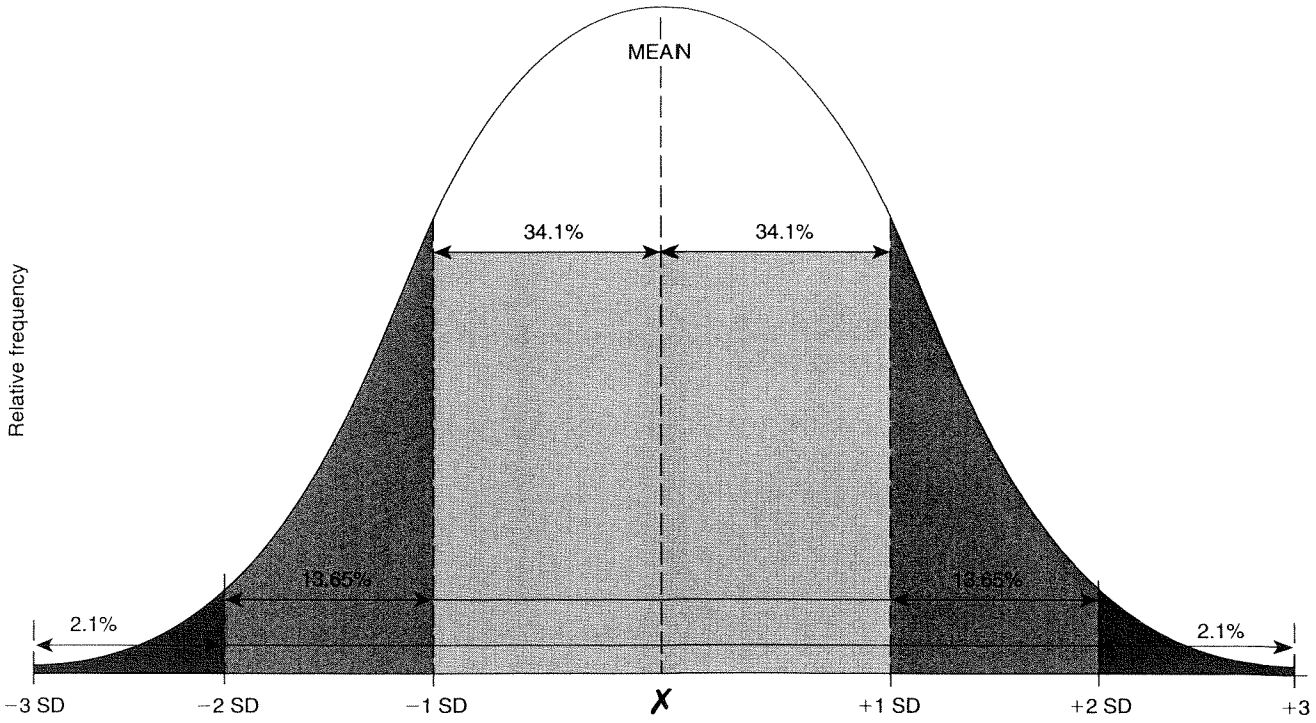

FIGURE 4-7

A normal distribution curve showing specific SD units.

specified SD units about the mean. The usual confidence limits used in the clinical laboratory are as follows:

68.2% confidence limits − the mean ± 1 SD.
68.2% equals the sum of 34.1% of the data on the left side of the mean and 34.1% on the right side.
95.5% confidence limits − the mean ± 2 SD.
95.5% is the total of 68.2% ($\bar{x} \pm 1$ SD) of the data plus two additional units of 13.65% from both sides of the mean.
99.7% confidence limits − the mean ± 3 SD.
99.7% is the sum of 95.5% of the data ($\bar{x} \pm 2$ SD) and two units of 2.1% from the final SD sections on the x-axis.

As you can see from Figure 4–8, the confidence limits combine equal areas from both sides of the mean that are bound by equal SD units. Even though the length of the SD units will change if the distribution changes from tight to broad, the area of the distribution curve defined by an SD unit will not change. Therefore, the percentage of data points contained within an area will not change with distribution changes.

If the data from repeat measurements do not exhibit a normal distribution, the SD and the mean ± SD confidence intervals cannot be used to describe the dispersion of data about the mean. This is because the SD is merely a mathematical description of a normal distribution. This is best illustrated in the following example:

If you were to calculate dispersion statistics for the data that produced the bimodal distribution curve in Figure 4–4, you would obtain the following results:

range = 96 − 106	mean = 99
variance = 7.7	3 ± SD = 8.4
SD = 2.8	mean ± 3 SD = 91 − 107

The calculated SD for these data is invalid because the confidence interval that should represent 99.7% of the data ($\bar{x} \pm 3$ SD) defines a range that is larger than the actual data range, an impossibility. Therefore, data must be determined to follow a normal distribution before the SD can be used to describe its dispersion about the mean.

Although the standard deviation gives an accurate mathematical picture of the spread of the data about the mean, it is often difficult to determine if the value for the SD is acceptable. Acceptability is judged by obtaining a relatively small number for the SD, which indicates that the data have a tight distribution about the mean.

Coefficient of variation (CV) is the standard deviation expressed as a percentage of the mean. The CV is calculated by dividing the standard deviation by the mean and multiplying by 100.

The formula for calculating the CV is:

$$\frac{\text{SD}}{\bar{x}} \times 100$$

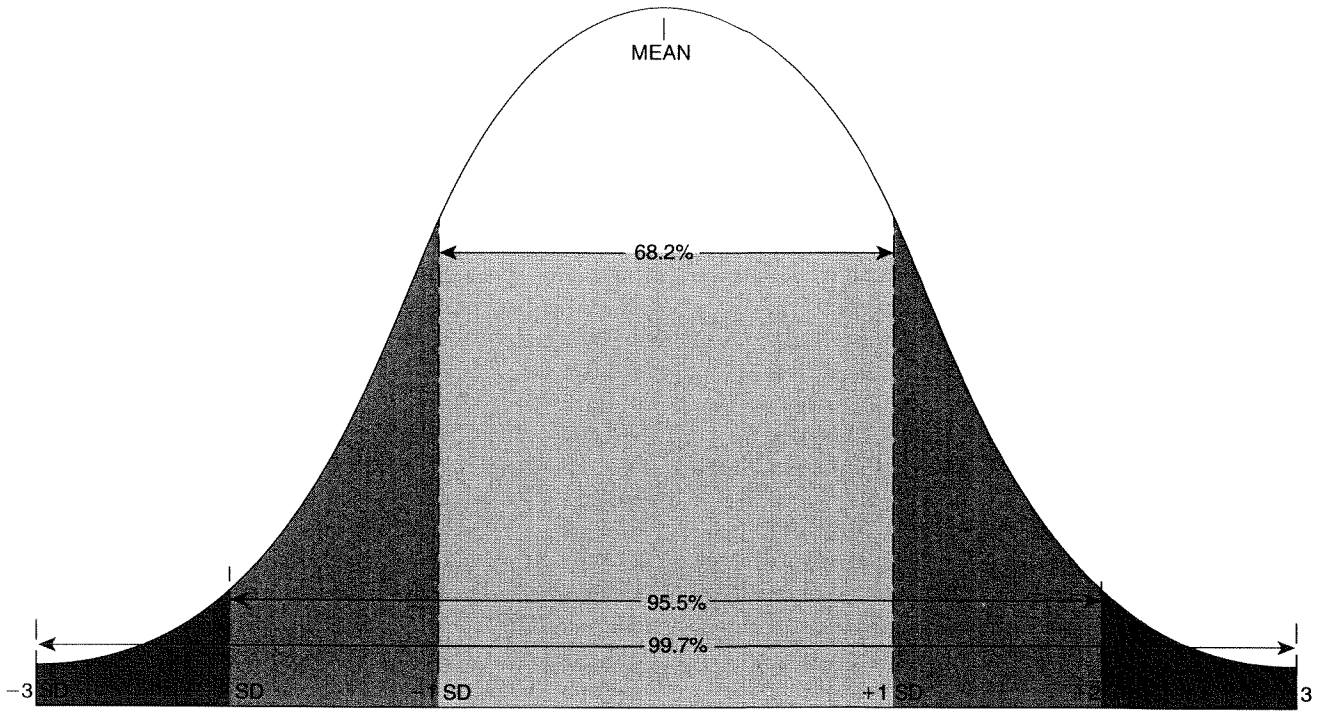

FIGURE 4-8

Confidence limits and SD units in a normal distribution curve.

Since the coefficient of variation is a percentage rather than a statistic with units as is the case with SD, it is easier to set criteria for acceptability. The usual limits for acceptability are that the CV on repeat laboratory measurements should be less than 5%.

The coefficient of variation for the data from Example 4–1 is 2.3% (2.3/99 × 100). This indicates that the size of the SD in relationship to the mean is acceptable and therefore the distribution is tight rather than broad about the mean.

Internal Quality Control Procedures

Once data from repeat measurements of a control are shown to have a normal distribution and the distribution is confirmed to be tight rather than broad about the mean, the control is ready to be put into use in the laboratory. The control is then assayed along with patient samples at intervals within the assay procedure that are defined by the method and by protocols within the laboratory. Control assay values are monitored against preestablished confidence intervals to verify accuracy and precision in measurement. **Accuracy** is the ability to obtain the established or "true" value for a sample, whereas **precision** is the ability to obtain the same value for repeat measurements of a sample.

These two definitions seem similar and, in fact, the terms are often used incorrectly. Accuracy and precision are very different concepts. Accuracy in measurement describes the correctness of a result, whereas precision in measurement describes the ability of the method to maintain the same value on repeat measurements and over time. Laboratory methods must be monitored for both accuracy and precision since it cannot be assumed that a method has both if only one or the other is confirmed. For example, a method may be precise in that repeat measurements are nearly the same, but the repeated value is not the true value, which therefore makes the method inaccurate.

In theory, it would be possible to have a method that is accurate but imprecise in that measurements do not repeat well but the overall average of the repeat values is near the true value. In practice, however, this rarely occurs. If a method is imprecise, it is usually also inaccurate.

A control monitoring system must give immediate information to the test operator (bench tech) who makes decisions about the quality of patient results. The system must also give periodic information to the laboratory supervisor who makes decisions about the overall performance of a method (precision and long-term accuracy of measurement).

LEVEY-JENNINGS CHARTS

One of the oldest methods for monitoring control values is a Levey-Jennings chart. A Levey-Jennings chart is a normal distribution curve (also referred to as a Gaussian distribution

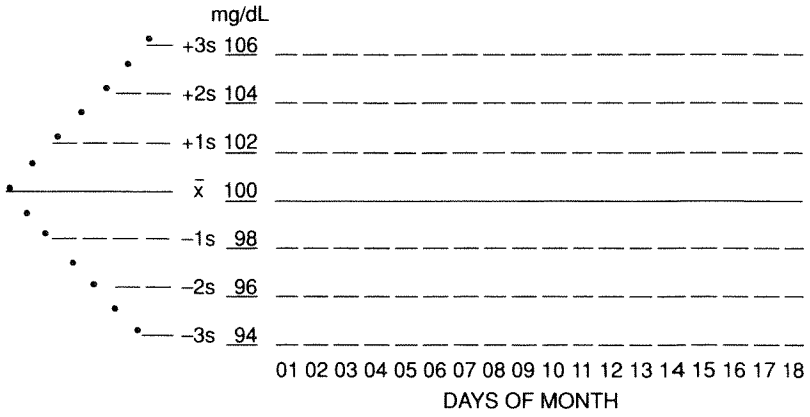

FIGURE 4-9

Levey-Jennings chart.

curve) lying on its side with specific points on the curve extended into a chart format designed for plotting values. This is shown graphically in Figure 4–9.

Example 4–4

A small laboratory is using a Levey-Jennings chart to record and monitor values for a midrange glucose control. The method being monitored is an automated glucose oxidase method that is run about 2 to 3 times per day. The laboratory assayed the control for 30 days to establish the target range. The mean for the data was 100 mg/dL and the SD was 2 mg/dL. The 95.5% confidence limit ($\bar{x} \pm 2$ SD) was 96 mg/dL to 104 mg/dL. The laboratory labels the chart and puts the control into regular use. Each control value is plotted at the time of assay and the values are monitored daily by the supervisor. The first 10 values obtained for the control after it was put into use are listed below. The resulting Levey-Jennings chart for these values is shown in Figure 4–10. Notice that the bottom of the chart in Figure 4–10 is designed to be labeled sequentially in the order that controls are run each day and over a number of days.

Date	Assay No.	Value
7-14	1	101 mg/dL
	2	100 mg/dL
	3	98 mg/dL
7-15	1	99 mg/dL
	2	102 mg/dL
7-16	1	103 mg/dL
	2	98 mg/dL
7-17	1	100 mg/dL
	2	101 mg/dL
	3	100 mg/dL

On a Levey-Jennings chart, the point that defines the mean on the Gaussian curve becomes a solid line located in the middle of the chart. Points on the curve corresponding to SD intervals become dashed lines located equidistant from the mean both above and below the midline. The left side of the chart is labeled with the mean and appropriate SD values at line intersection points. The base of the chart is labeled from left to right in time units that reflect the sequence of control measurements. The time units could be days, number of sequential values over several days, or numbers of values within one day, but should cover a period sufficient to display at least 30 sequential assay results.

The values on the left side of a Levey-Jennings chart are the values obtained from the repeat control measurements that were performed for the purpose of establishing target ranges. Once data from these measurements are collected and shown to be normally distributed, confidence intervals are calculated and mean and SD points of intersection are labeled on the chart. Once the chart is labeled, it is then used to monitor control values.

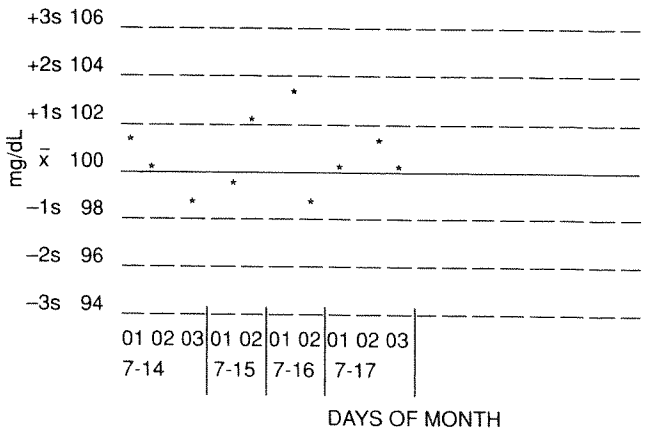

FIGURE 4-10

A Levey-Jennings chart using a midrange control serum with an automated glucose oxidase procedure.

In the simplest Levey-Jennings QC monitoring system, immediate decisions about the quality of patient results are based on the degree to which control values remain within a preestablished limit. If a control value is within this limit, the assumption is made that results on patient samples run concurrent with the control are correct. For example, a common acceptable limit for a control might be that all values must be within the 95.5% confidence interval, which is shown on a Levey-Jennings chart as the area covered by the mean ± 2 SD. In other words, current assay values for the control must be essentially the same as 95.5% of the data obtained on the control during the period that the target range was established. Using this control rule criteria, a control value is considered acceptable if it lies within the shaded area shown in Figure 4–11.

Precision and long-term accuracy of a method are confirmed on a Levey-Jennings chart by control values remaining clustered about the mean with little variation in the upward or the downward direction as shown in Figure 4–12. Note also in Figure 4–12 that approximately the same number of points lie above the mean as below the mean.

Imprecision in measurement is indicated by a large amount of scatter about the mean and usually an uneven distribution above and below the mean. The Levey-Jennings chart in Figure 4–13 shows control values that demonstrate imprecision in measurement.

Imprecision is most often caused by technique errors such as variability in pipetting or inattention to detail by the operator. Imprecision results in an increase in the SD or a broadening of the distribution about the mean.

Inaccuracy that occurs over time is also detected by a Levey-Jennings chart. Changes in long-term accuracy are subtle changes that occur in a method and often go unnoticed by the operator who is making immediate decisions about patient results. This type of inaccuracy occurs because there has been a change in the measurement process that is not large enough to be noticed immediately but that can affect patient results. Long-term inaccuracy is indicated on a Levey-Jennings chart by either a trend or a shift.

A **trend** is a gradual change in the mean that proceeds in one direction.

Figure 4–14 shows a Levey-Jennings chart of control values that demonstrate a trend occurring from value 9 through value 20. A trend is usually caused by gradual changes in the method of assay such as deterioration of reagents and standards or deterioration in instrument performance. Values may still be clustered about the mean but the mean itself is gradually trending toward either higher or lower values.

A **shift** is an abrupt change in the mean that becomes continuous.

The Levey-Jennings chart in Figure 4–15 shows a shift in control values that occurred at value 9. A shift is caused by the introduction of something new into the assay procedure. The usual causes of shifts are new lots of standards and reagents or a slight malfunction in an instrument that results in an

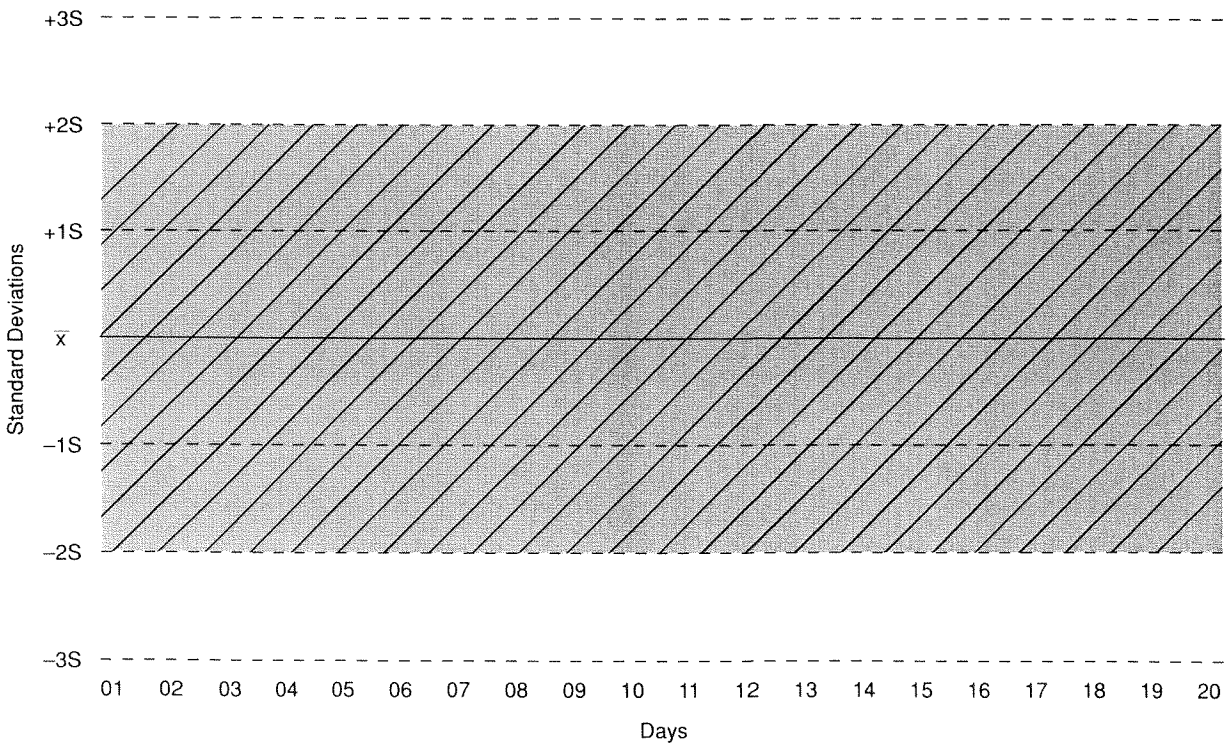

FIGURE 4-11

Levey-Jennings chart.

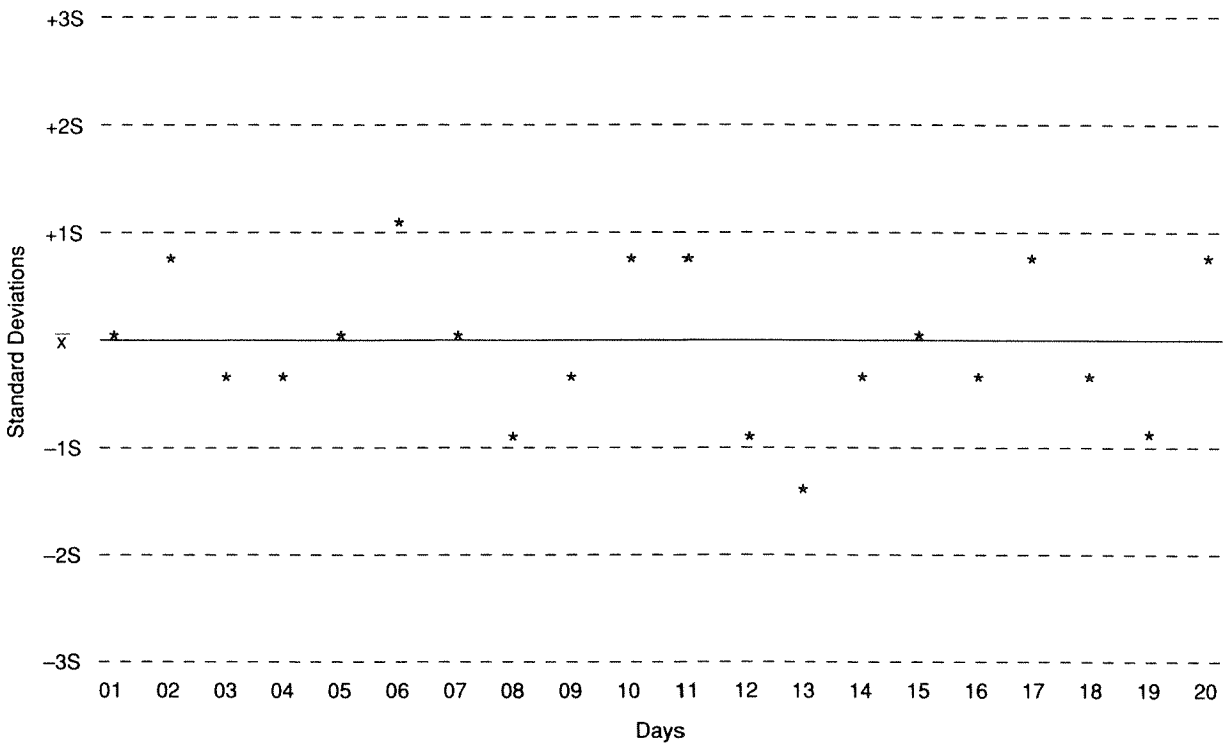

FIGURE 4-12

Levey-Jennings chart demonstrating precision.

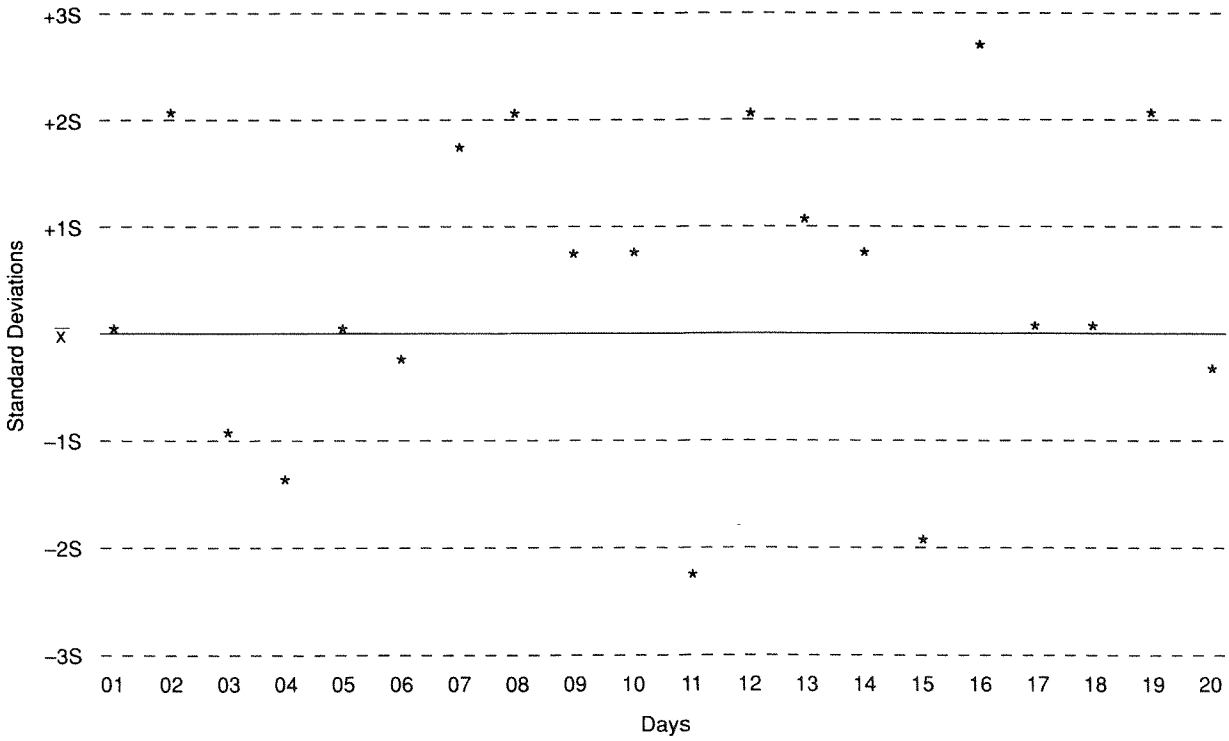

FIGURE 4-13

Levey-Jennings chart that demonstrates imprecision.

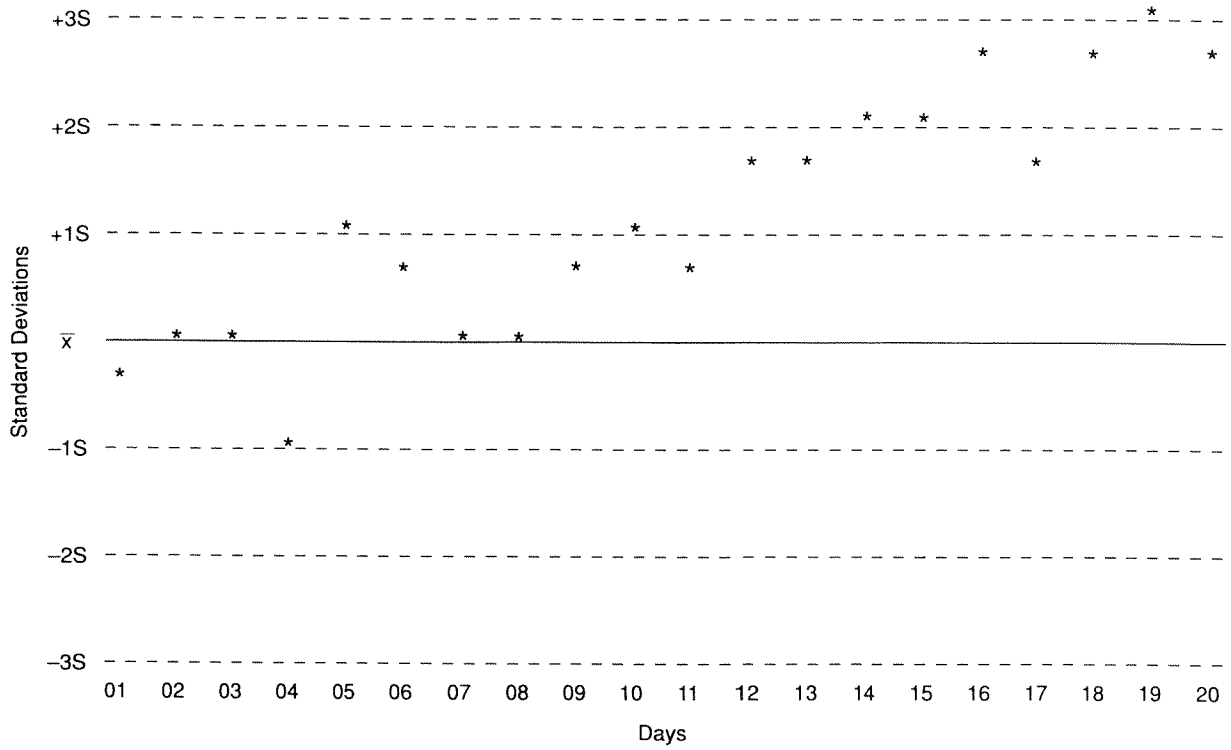

FIGURE 4-14

Levey-Jennings chart that demonstrates a trend.

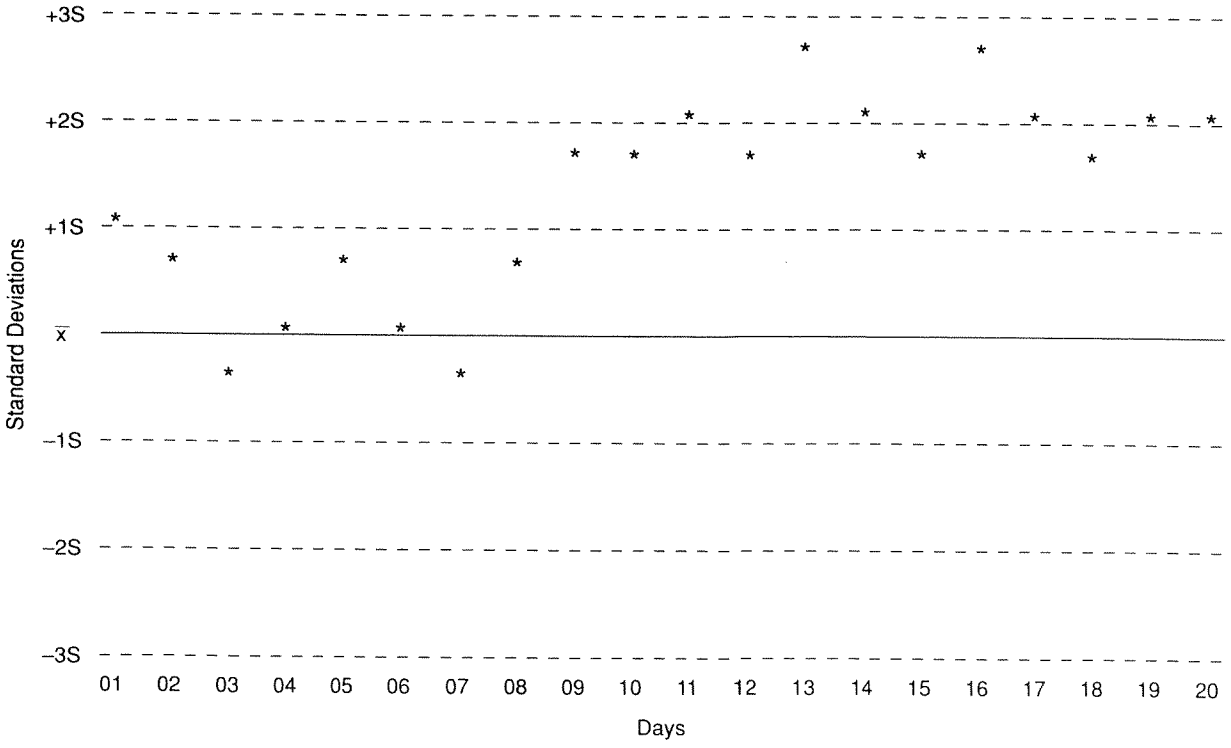

FIGURE 4-15

Levey-Jennings chart that demonstrates a shift.

immediate and somewhat permanent change in performance. The exact time the new solution was introduced or the change occurred can be pinpointed on a Levey-Jennings chart as the point at which the mean shifted to either a higher or a lower value.

The most obvious advantage of a Levey-Jennings chart is that it gives a good visual representation of precision and long-term accuracy and is easy to interpret. The disadvantage of a Levey-Jennings chart is the amount of time required to maintain chart data. For effective use, control values must be plotted at the time of assay and in the order of measurement. In addition, separate charts are required for each assay and each level of control. Computers can be very useful tools in maintaining and analyzing the large of amount of data generated.

Error Detection

Since the goal of QC is to monitor processes and detect problems, it may be useful to think of the QC process as an alarm system. The system collects data (control values) about the process, and if these values exceed some preestablished value, an alarm is sounded. Similar to the temperature monitors on many laboratory refrigerators or freezers, they measure temperature and sound an audible alarm if the temperature goes too high or too low. The values for these "too high" and "too low" temperatures are usually set by the user. But how high is too high? Is it okay for the 4°C refrigerator to reach +10°C? Or to go down to −20°C and freeze the items inside? Someone had to do some research, create a plan, and implement the high and low temperature limits for the alarm system. Quality control is just the same. To have a good QC alarm, one has to define what is the acceptable range of error and then select tools that will be capable of detecting that error.

Control monitoring and error detection can be characterized by two key factors: (1) the probability of error detection (P_{ed}), and (2) the probability of false rejection (P_{fr}).

The probability of error detection describes the ability (probability, since this is a statistical analysis) of the detector to detect an error that is beyond the stable imprecision of the instrument. Ideally, our detector would detect 100% of the errors ($P_{ed} = 1.00$). In the clinical laboratory, a practical objective is 90% detection of errors.[3]

The probability of false rejection describes the probability of an alarm when no error exists. Ideally, the detector would not sound an alarm if nothing is wrong ($P_{fr} = 0.00$). However, think of your home smoke detector. How many times has it sounded when no dangerous fire existed? This is an example of a false rejection. Detectors can be designed to differentiate the smoke from your burned toast and that of your burning draperies, but it is practically difficult in terms of cost. Thus, smoke-detector manufacturers have to balance the sensitivity of their device to detect any smoke with the possible severity of the outcome (your death in a real fire). The device needs to have a very high probability of error detection. Manufacturers thus have to accept a number of false rejection alarms (burned toast sets off the alarm) to improve the ability to detect all fires (a 100% probability of error detection). Laboratory quality control faces a similar situation. The QC process needs to be able to detect errors before they cause harm to patients but a laboratory cannot afford (financially or in terms of time spent) to have a process that is so sensitive to changes that there are large numbers of false rejection and results in continually trying to find a problem when none exists. Careful planning (quality planning) is required for proper selection of control processes and rules that are sensitive enough to prevent harm but avoid large numbers of false rejections.

QUALITY PLANNING FOR QUALITY CONTROL

As stated previously, the first step in quality is the definition of quality goals, objectives, and the establishment of quality requirements. Analytical quality results from a complex relationship between the imprecision and inaccuracy of the method, normal physiologic variation in the analyte, changes in the specimen after it is collected, or even differences in the sample that result from collection procedures. Each analyte has a different set of the listed characteristics and, as such, the quality of measurement cannot be measured with the same rule system for each. If you haphazardly select rules, you may have a high false-rejection rate or have a low error-detection rate and, more important, you will not have any valid assurance that you are getting quality results.

Fortunately, the laboratory technologist does not need to be an expert in the establishment of quality goals and requirements. There are a number of tools available from professional organizations, websites, and other resources. Guidelines created by regulatory agencies, standards creating organizations (e.g., National Committee for Clinical Laboratory Standards, NCCLS) can be found in publications, texts, and on the WWW.[4,5] Essentially, the process requires three steps. The laboratory (1) defines the quality required by the test, (2) measures the imprecision and inaccuracy of the method, (3) selects rules based on their probability of error detection and false rejection.

Resources for defining quality requirements include the allowable error as defined by the Proficiency Testing limits under the Clinical Laboratory Improvement Act (CLIA), medically relevant limits found in the literature, or in consultation with physicians.[4–8] These limits describe how far apart the measures on the same specimen can be before the difference changes clinical decisions. For example, under CLIA, acceptable variation in sodium is the target value ± 4 mmol/L. Thus, if you run a sodium analysis and get 140 mmol/L, then rerun that same sample and get 142 mmol/L, that is acceptable quality. If you got 148 mmol/L on the repeat test, that would not be acceptable quality because the clinician would see that as a significant change and think something has changed in the patient.

The measures of imprecision and inaccuracy are conducted as part of the method evaluation when a test is being installed. After a test has been in use, the bias (inaccuracy) can be determined from peer comparison programs or proficiency testing survey results. The imprecision can be determined from the daily QC values. If the measures of imprecision and inac-

curacy fall outside what you determined to be acceptable quality, your method is not acceptable and needs to be improved or another method selected.

Once you have determined the quality required and the performance (inaccuracy, imprecision) of your method, then you need to select control rules and determine the frequency that they need to be run so you will know when the quality of results is declining. There are a number of ways to select these rules to maximize error detection and minimize false rejections. Readers are referred to the multitude of resources for rule selection found in journals, quality control texts, and on the WWW. The Westgard QC website provides numerous tools to assist with this process (www.westgard.com).

Multirule Systems

Although simple control rules may provide the required error detection and avoid significant false rejection (as determined during the quality planning process), it is more common to see a multirule system in place. These systems use multiple rules that define specific limits for control values. If a control value violates a rule by exceeding the limits, an error in measurement has been detected and the results on patient samples assayed concurrently with the control cannot be released to the physician. A common set of laboratory QC multirules is Westgard's multirule system.[9] Control rules are usually described using abbreviations such as 1-2S or 4-1x, where the first number represents the number of control values that have been observed and the second and third characters describe the control limit. For example, 1-2S means that one control value was more than two standard deviations away from the mean value. The 4-1x notation indicates that four control values were more than one standard deviation away from the mean. Other texts and papers may use slightly different notations (e.g., 1_{2s} or 1:2s). The original Westgard rules and the definitions for each are presented below. Levey-Jennings charts giving visual representations of the violations for each rule are presented in the accompanying figures.

1-2S—One control value exceeds the mean by more than 2 SD but less than 3 SD. The control can exceed the mean in either the upward or the downward direction. This rule is demonstrated in Figure 4–16.

1-3S—One control value exceeds the mean by more than 3 SD in either the upward or the downward direction. The violation of this rule is shown in Figure 4–17.

2-2S—Two consecutive control values exceed the mean by more than 2 SD but less than 3 SD. These two control values must be consecutive and they must lie in the same

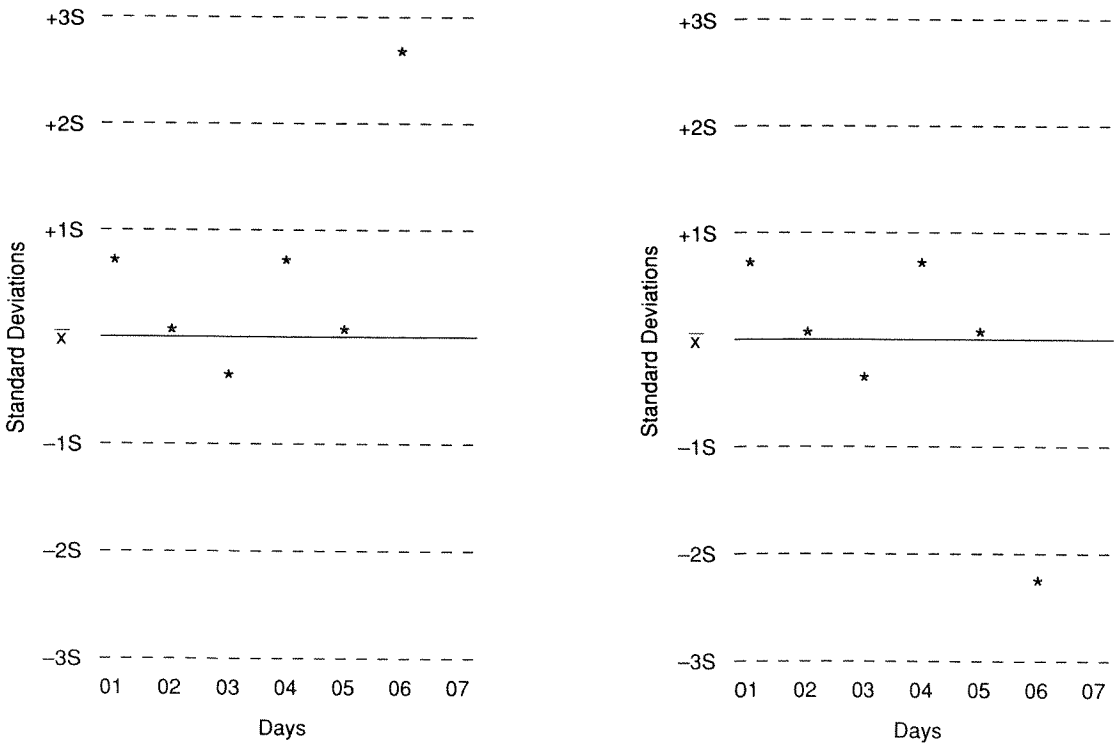

FIGURE 4-16

Levey-Jennings chart representing the 1-2S multirule system. *Left*, a control value exceeds 1-2S in the upward direction on day 6. *Right*, a control value exceeds 1-2S in the downward direction on day 6.

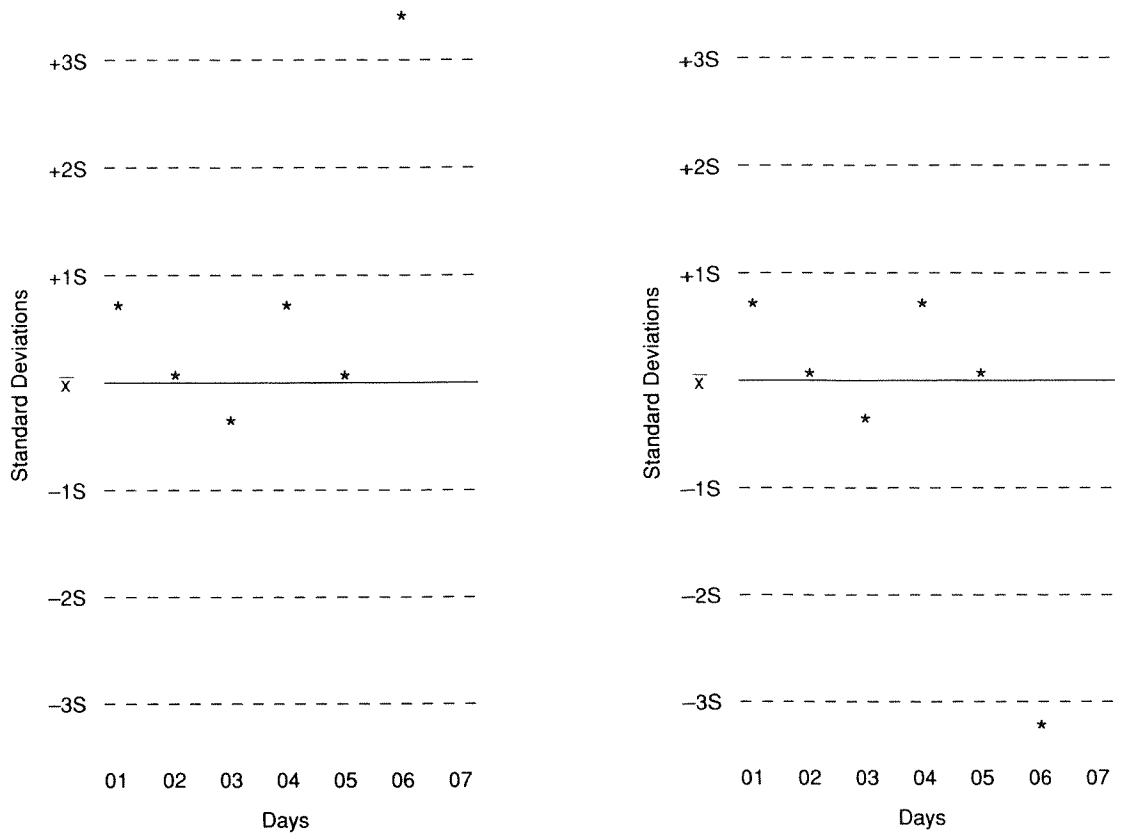

FIGURE 4-17

Levey-Jennings chart representing the 1-3S multirule system. *Left*, a control value exceeds 1-3S in the upward direction on day 6. *Right*, a control value exceeds 1-3S in the downward direction on day 6.

direction from the mean. This rule violation is shown in Figure 4–18.

R-4S—The difference between two consecutive controls is greater than 4 SD. These two consecutive control measurements have assay values in opposite directions from each other and the difference between the two spans at least 4 SD. The violation of this rule is shown in Figure 4–19.

4-1S—Four consecutive control values exceed the mean by more than 1 SD. These four controls must be consecutive and lie in the same direction from the mean. The violation of this rule is shown in Figure 4–20.

10x—Ten consecutive control values exceed the mean in the same direction. This rule violation is shown in Figure 4–21.

In the original Westgard multirule system, the 1-2S rule is a flag that indicates a possible change in long-term accuracy or precision. If a control value exceeds 2 SD in either direction but is less than 3 SD, the remaining rules are applied to the data. If there are no violations of the remaining rules, the control values are considered acceptable and patient assay results are released to the physician. Thus the 1-2S flag does not always indicate error in measurement.

The justification for using the 1-2S rule as a flag and not a violation is based on the definition for the 95.5% confidence limit of a control's target range data. By definition, the 95.5% confidence limit ($\bar{x} \pm 2$ SD) contains 95.5% of the data points obtained while establishing the target range for the control. Conversely, 4.5% of the control data, although representing very real values, are outside these limits. There is no requirement to use rules as flags (or warning rules). The use of the 1-2S rule as a flag is often designed to save time and effort when interpreting control data manually. When computers are used to interpret control data a flag or warning rule may not be utilized.[10]

When a control is put into use and run alongside patient samples, it is expected that assay values will duplicate those obtained on the control during the target range period. As mentioned previously, a common rule for acceptance of control data is that all control values must be within the 95.5% confidence limit. This criteria would mean that 1 in every 20, or 4.5%, of the values would be out of this range due to chance alone even though the control may be duplicating values from the target period. In fact, if the control is rerun, chances are the assay value will now be within the 95.5% confidence limits. NOTE: It is not appropriate to simply rerun a control that exceeds the 1-2S limit. You must follow the defined control rules and base your decision on that result. If you exceed the

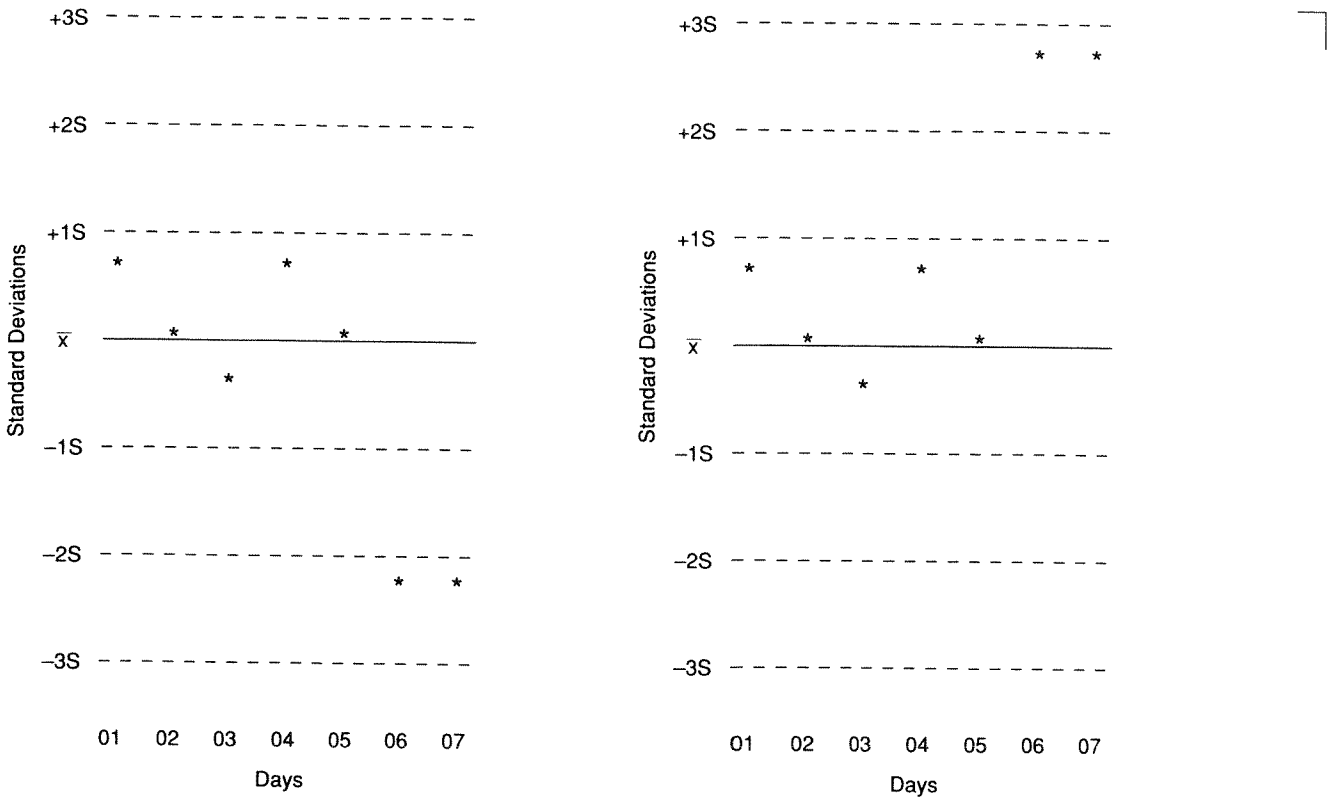

FIGURE 4-18

Levey-Jennings chart representing the 2-2S multirule system. *Left,* two consecutive control values on day 6 and on day 7 violate the 2-2S rule in the downward direction. *Right,* two consecutive control values on day 6 and on day 7 violate the 2-2S rule in the upward direction.

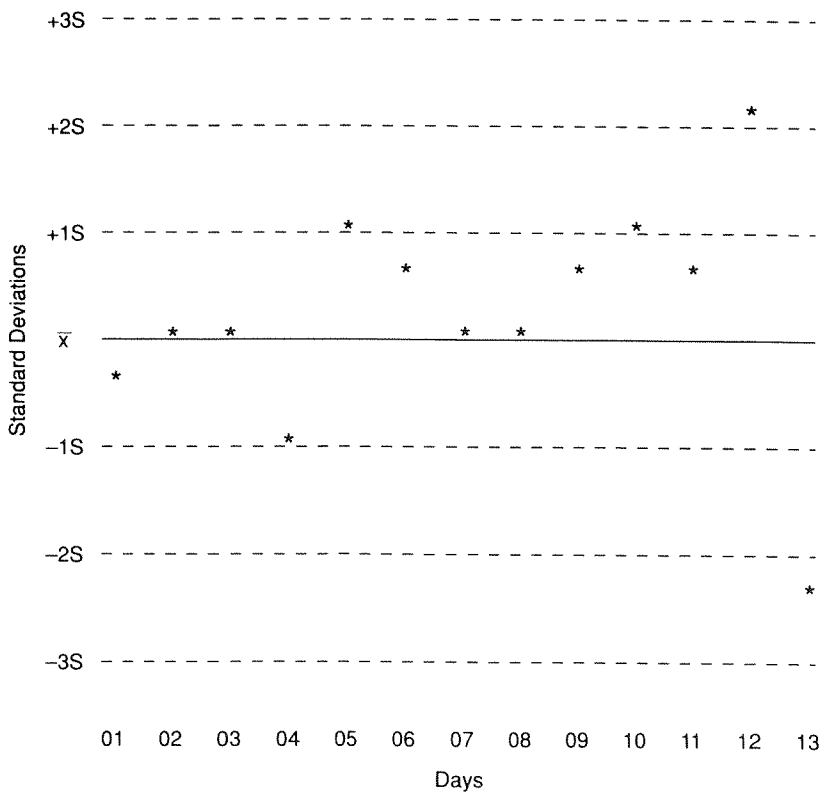

FIGURE 4-19

Levey-Jennings chart representing the R-4S multirule system. The difference between two consecutive control values (days 12 and 13) exceeds 4 SD.

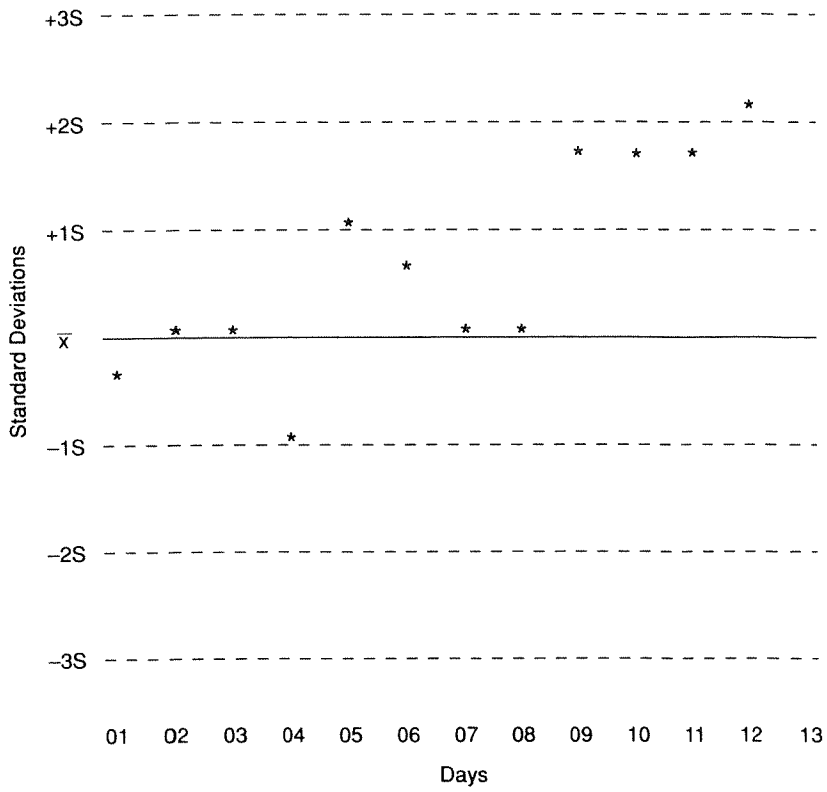

FIGURE 4-20

Levey-Jennings chart representing the 4-1S multirule system. Four consecutive controls (days 9 through 12) violate the 4-1S rule in the upward direction.

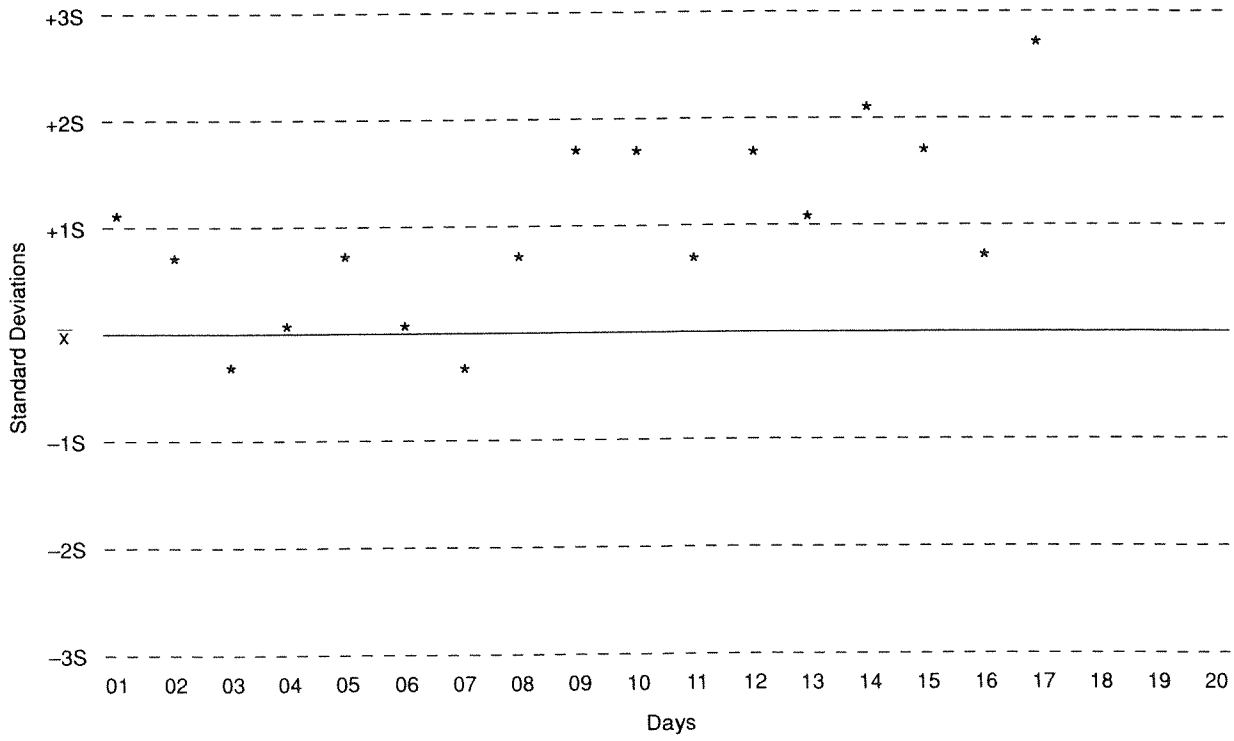

FIGURE 4-21

Levey-Jennings chart representing the 10x multirule system. Ten consecutive control values (days 8 through 17) violate the 10x rule in the upward direction.

Example 4–5

A laboratory is using the Westgard multirule system to monitor control values on an automated instrument. The following results are twelve consecutive assay values for the calcium control and are in units of mmol/L. Values flagged by the 1-2S rule are indicated and any rule violations are listed after the flag.

Target values for control: mean = 3.3 mmol/L
and SD = 0.1 mmol/L

Mean ± 1 SD range = 3.2 to 3.4 mmol/L
Mean ± 2 SD range = 3.1 to 3.5 mmol/L

Consecutive Order of Controls	Value	Flag	Westgard Rule Violated
1	3.2		
2	3.3		
3	3.4		
4	3.2		
5	3.1	1-2S	none
6	3.3		
7	3.4		
8	3.5		
9	3.6	1-2S	none
10	3.6	1-2S	2-2S

1-2S limit (a warning in the Westgard multirule system) but no other control rules are violated, you have valid assurance that nothing is wrong and patient results can be reported. If you rerun the control material, you are wasting reagent, time, and control material.

The original Westgard system attempts to beat those 1 in 20 odds by using the 1-2S rule as a flag and not as a violation. If, in addition to being outside the 95.5% confidence limit, a control value also violates a second rule, we can be more confident that there has been a change in assay performance since the establishment of the target range.

Before presenting a few exercises on the use of multirules, it is necessary to review a few points about the application of control rules:

1. Control values must exceed the limits to be considered a violation. For example, if the 95.5% confidence limits for a glucose control were 95 to 100 mg/dL and one control assay value was 100 mg/dL, it would not exceed the $\bar{x} \pm 2$ SD limit. A value of 101, on the other hand, would exceed the limit.
2. The value of the control that was the 1-2S trigger is included in the count of consecutive controls for applying rules. This is most easily demonstrated by the 2-2S rule. As can be seen from Figure 4–18, the value on day 6 was a flag that triggered the other rules but none were violated; therefore, patient values were accepted and results were sent to the physician. The value on day 7 was again a flag, but now two flags have occurred consecutively. The flag on day 7 is included in the count when applying the 2-2S rule and the rule has now been violated.
3. The R-4S, 4-1S, and 10x rules are not applied unless a 1-2S flag has occurred to initiate their application (the 2-2S rule contains the flag as a part of the violation and the 1-3S rule is always a violation even without a flag). In Figure 4–21, the values on days 12, 13, 14, and 15 all exceed $\bar{x} \pm 1$ SD. However, the 4-1S rule has not been violated because a 1-2S flag did not occur to initiate the application of the other rules. On day 17, a 1-2S flag did occur and when the rules were applied, the 10x rule was violated. The 4-1S rule was not violated because on day 16, the value of the control did not exceed the mean plus 1 SD.

Example 4–5 demonstrates the use of the Westgard system for monitoring control data.

Example 4–6

The following data are from a laboratory that assays two levels of controls for potassium. The values for the two levels are given in the order in which they were assayed.

Target values (high control)

Mean = 6.6 mmol/L
$\bar{x} \pm 1$ SD = 6.5 to 6.7 mmol/L
$\bar{x} \pm 2$ SD = 6.4 to 6.8 mmol/L

Target values (low control)

Mean = 3.5 mmol/L
$\bar{x} \pm 1$ SD = 3.4 to 3.6 mmol/L
$\bar{x} \pm 2$ SD = 3.3 to 3.7 mmol/L

Consecutive Order of Controls	Value	Flag	Westgard Rule Violated
1 low	3.6		
2 high	6.7		
3 low	3.3		
4 high	6.4		
5 low	3.7		
6 high	6.3	1-2S	none
7 low	3.5		
8 high	6.6		
9 low	3.4		
10 high	6.5		
11 low	3.7		
12 high	6.8		
13 low	3.7		
14 high	6.9	1-2S	4-1S across controls

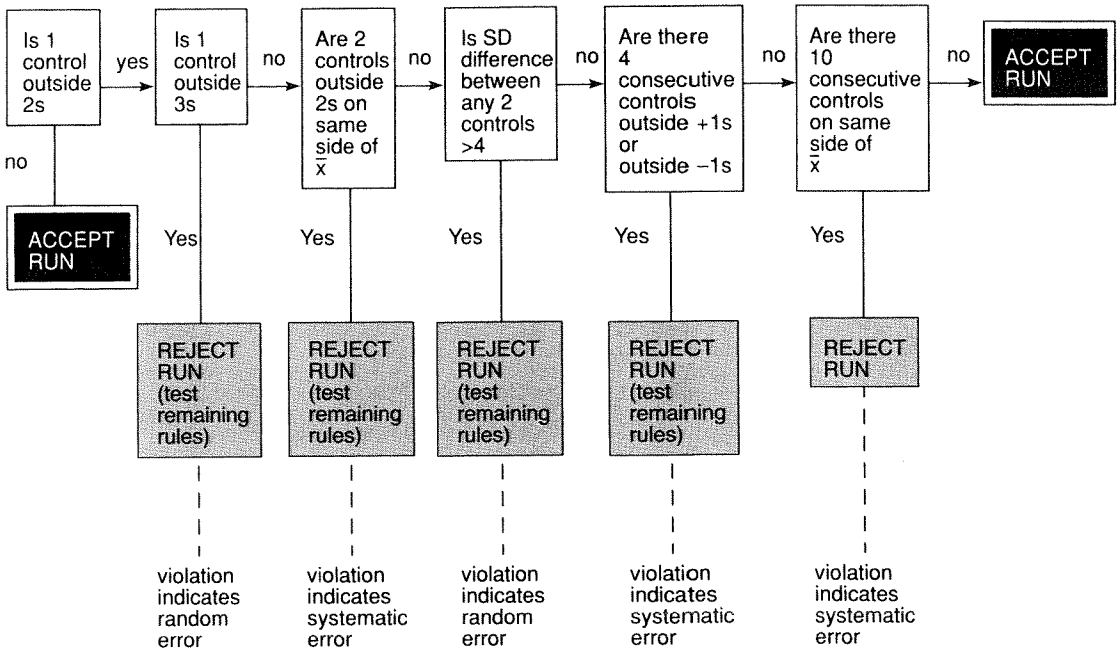

FIGURE 4-22

The Westgard multirule system.[11]

The greatest advantage to any multirule system is the power of the rules for distinguishing between random and systematic error. **Random error** is the error that occurs without any real pattern and **systematic error** is the error that is continuous and affects all results equally.

On a Levey-Jennings chart, random error is indicated by a control value that is significantly different from the other values. Systematic error on a Levey-Jennings chart is indicated by a trend or a shift. Systematic errors are errors that require investigation to identify their causes. Random errors can usually be considered a one-time event and patient samples and controls can be rerun with success. The exception to this is that a consistent random error problem requires investigation because it could indicate a change in precision.

In the Westgard system, specific rules indicate random error and other rules indicate systematic error. As shown in Figure 4–22, the 1-3S rule and R-4S rule are rules whose violation indicates the presence of a random error. The 2-2S, 4-1S, and 10x rules are rules whose violation indicates a systematic error.

In addition to distinguishing between random and systematic error, the Westgard system also indicated the direction that the investigation of a systematic error should take. This ability to indicate the direction for investigation of error is realized when a specific process for applying the rules is followed:

1. The rules are applied across controls. This is illustrated for a two-control system in Example 4–6. In this example, the value on day 14 exceeded the mean plus 2 SD and was thus a flag that initiated the application of the rules to previous control data. Values on days 14, 13, 12, and 11 are all above the mean plus 1 SD. The values are consecutive and lie in the same direction from the mean. Therefore, Westgard rules indicate a systematic error has occurred. The rules can be applied across control pools for systems containing three or more control pools in a similar manner.
2. The rules are applied across runs. This may require a system for retrieving data such as a computer monitoring system. The 4-1S rule and the 10x rule are the most powerful rules for application across consecutive runs or days.
3. Once a rejection is identified, the operator continues to apply the rules until all have been tested. This is illustrated in Example 4–7. In this example, the operator did not stop after the first rule rejection and ultimately identified three rule rejections. If the process had been terminated after the 4-1S rejection within control II, the full extent of the error would not have been realized. In this example, the systematic error extended throughout the range represented by both controls. A violation within a single control, on the other hand, might be due merely to a loss of calibration at that end of the linear range as opposed to a complete instrument malfunction.

Example 4–7

The following data are from a laboratory which assays two levels of controls for an analyte. The values for the two levels are given in the order in which they were assayed.

Target Values (control I)

Mean = 16.1 mmol/L
$\bar{x} \pm 1$ SD = 15.8 to 16.4 mmol/L
$\bar{x} \pm 2$ SD = 15.5 to 16.7 mmol/L

Target Values (control II)

Mean = 25.6 mmol/L
$\bar{x} \pm 1$ SD = 25.2 to 26.0 mmol/L
$\bar{x} \pm 2$ SD = 24.8 to 26.4 mmol/L

Consecutive Order of Controls	Value	Flag	Westgard Rule Violated
1 I	16.2		
2 II	25.8		
3 I	16.0		
4 II	25.4		
5 I	16.2		
6 II	26.1		
7 I	16.5		
8 II	26.2		
9 I	16.3		
10 II	26.3		
11 I	16.6		
12 II	26.3		
13 I	16.7		
14 II	26.5	1-2S	4-1S within control II 4-1S across controls 10 x across controls

External Quality Control

The QC procedures addressed previously in this chapter are referred to as internal quality control. A laboratory QC program is not complete unless it also includes an external component. External quality control is a process by which a laboratory uses an outside unbiased source to verify the quality of patient results.

The most common type of external quality control involves testing biological samples submitted to a laboratory by an outside agency. The laboratory assays the unknown samples "blind" just as though they were actual patient samples and sends the results back to the agency. The laboratory's accuracy of measurement is also compared with other laboratories using similar methods. This method of external QC is called *proficiency testing* and is often tied to requirements for laboratory accreditation.

Proficiency testing is administered by a number of companies including the American Proficiency Institute, American Association of Bioanalysts, and the College of American Pathologists. These companies must be approved by the Health Care Financing Administration (HCFA) and successful participation in a recognized proficiency testing program is required for proper compliance with CLIA '88. In addition to meeting HCFA requirements, some accrediting agencies will have more stringent or additional proficiency testing requirements to maintain accreditation.

The acceptable limits for passing proficiency events are defined by HCFA.[4] Qualitative tests are graded using the result that 90% of participants provided. For example, if 100 laboratories participate in a pregnancy testing survey and 92 report the sample as positive, then the correct answer is positive (and the 8 laboratories reporting negative will not pass). Immunohematology results are an exception to this rule and are based on a 95% consensus.

Quantitative results are graded against a target value ± a fixed amount. The fixed amount can be expressed as a percentage, a number of standard deviations, or a quantity. For example, HCFA requires that chloride results be target value ± 5%, whereas total calcium is graded against target value ± 1.0 mg/dL and TSH is graded against target value ± 3 SD. The target value is the mean value obtained by a peer group using the same instrument or reagent as your laboratory. A peer group must have at least 10 laboratories participating to be valid. Additionally, like qualitative results, only those analytes having a 90% consensus are graded.

The statistical term standard deviation interval (SDI) quantitates the number of SD units a single result differs from the group mean. The SDI is used by proficiency organizations to define the amount a lab result may differ from the average of the results from a group of peers. This difference is then adjusted for the average difference (standard deviation) of the group. The calculation of an SDI is as follows:

$$\text{SDI} = \frac{\text{Lab result} - \text{Peer group mean}}{\text{Peer group SD}}$$

Example 4–8 shows proficiency survey results using both fixed criteria and standard deviation intervals to report results. As can be seen in Example 4–8, laboratory A's potassium result on the proficiency sample is outside the acceptable limits. The SDI for their result on the BUN proficiency sample is acceptable, though it is close to the limit.

Since proficiency testing is closely tied to accreditation, the original intent of verifying accuracy in measurement by testing unknown samples blind is often overshadowed by the concern that a wrong answer may affect accreditation. For that reason, in addition to proficiency testing, many laboratories also subscribe to a method for external QC that does not carry

penalties for incorrect answers. As an example, some lab groups have established their own regional proficiency testing programs in which samples are sent to participating laboratories by one of the larger laboratories in the group. The results and comparisons are reported in the same manner as the agencies involved in proficiency testing. This regional approach attaches no penalty to wrong answers and the original concept of self-evaluation by peers is retained.

Example 4–8

On a proficiency survey, fixed criteria of ± 0.2 mmol/L were used for potassium evaluation. For BUN, limits of acceptability were ± 2 SDI from peer group results. Laboratory A's proficiency report is as follows:

	Lab A Result	Target Value	Peer Group Mean	Peer Group SD
Potassium	4.0	4.3		
BUN	24		21	1.5

The acceptability limits for potassium are calculated as follows:

Target value ± allowable concentration deviation
(4.3 + 0.2 = 4.5; 4.3 − 0.2 = 4.1) = 4.1 to 4.5

The SDI interval for Laboratory A's BUN result is as follows:

$$\text{SDI} = \frac{\text{Lab result} - \text{peer group mean}}{\text{Peer group SD}}$$

$$\frac{24 - 21}{1.5} = 2 \text{ SDI}$$

Some laboratories subscribe to external QC programs provided by companies who supply daily control materials to laboratories. The laboratory uses the control materials for daily internal QC procedures and sends statistical or raw data on the daily control values to the company at the end of each month. The company provides the laboratory with information on the amount of long-term inaccuracy and imprecision that may be occurring within the laboratory. The company also provides the laboratory with information on the closeness of their results on the control pools to peer laboratories using the same control materials. This information is provided in much the same format as provided by the proficiency testing groups.

SUMMARY

Quality assurance is the result of processes by which a laboratory defines, measures, and improves quality in laboratory results by closely monitoring the preanalytical, analytical, and postanalytical stages of laboratory testing. The process can be defined with a total quality management approach to quality that includes quality planning, processes, assessment, control, and improvement.

Quality control is a section of the analytical stage of quality assurance and is the process of monitoring results from control samples to verify quality in results from patient samples run alongside the controls. QC processes have both an internal and an external component.

The first step in an internal QC process is to establish the target range for the control sample. The target range is set by assaying the control repeatedly and verifying the existence of a point of central tendency and a tight distribution in the resulting data set. Once the range has been established, control results are monitored over time to detect any significant change in values that might indicate variability in method performance and a possible imprecision or inaccuracy in patient results. Imprecision in results is known as random error, and inaccuracy is termed systematic error. The monitoring of control results to detect random and systematic error is done by visual methods such as Levey-Jennings charts or by control rule methods that rely on rule violations to indicate that an error has occurred. The selection of rules to achieve error detection while reducing false rejection requires a thorough planning process that determines the quality required and is based on the known characteristics (bias, imprecision) of the method.

External QC is the process of assaying unknown samples from an outside agency and thus verifying accuracy of testing by comparison to an established value for the outside sample or comparison to the average result obtained on the sample by laboratories within a peer group. The most common external quality control systems are proficiency testing programs.

REFERENCES

1. Westgard JL, Burnett RW, Bowers GN: Quality management science in clinical chemistry: A dynamic framework for continuous improvement of quality. *Clin Chem* 36:1712–1716, 1990.
2. Westgard JO: Assuring quality through total quality management. Available at: http://www.westgard.com/essay4.htm
3. Westgard JO: QC—The chances of rejection. Available at: http://www.westgard.com/lesson15.htm#alarm
4. U.S. Department of Health and Human Services: Clinical laboratory improvement amendments of 1988; final rules and notice. 42CFR Part 493. Federal Register 57:7188–7288, 1992.
5. Guideline: Internal quality control testing: Principles and definitions. Villanova, PA, National Committee for Clinical Laboratory Standards (C24-T), 1987.
6. Skendzel LP, Barnett RN, Platt R: Medically useful cri-

teria for analytic performance of laboratory tests. Am J Clin Pathol 83:200–205, 1985.
7. Fraser CG: Biological variation in clinical chemistry. An update: collated data, 1988–1991. Arch Pathol Lab Med 116:916–23, 1992.
8. Westgard JO, Seehafer JJ, Barry PL: Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. Clin Chem 40;1909–1914, 1994.
9. Westgard JO, Barry PL, Hunt MR, Groth T: A multirule shewhart chart for quality control in clinical chemistry. Clin Chem 27:493–501, 1981.
10. Westgard JO: QC—The multirule interpretation. Available at: http://www.westgard.com/lesson18.htm
11. Bakes-Martin R, Romph P, Myers K: Quality control pool evaluation. *Basic Quality Assurance for the Medical Laboratory, LAP LM-5—A Self Study Course.* Denver, CO Association for Continuing Medical Laboratory Education, March 1988.